Journal of
Social and Economic Statistics

*Vol. 12, No. 2, Winter 2023*
**DOI**: 10.2478/jses-2023-0008



# *MAPPING THE PARAMETER SPACE OF SIMULATED LOTTERIES* *

**István GERE**[a*]

[a)] *Department of Physics, Babeş-Bolyai University, Cluj-Napoca, 400347, Romania*

**Abstract**

*This study presents an econophysics based approach to the study of lotteries. By treating lotteries as complex systems we analyse the guiding dynamics of real-world lotteries. We found that the growth of the jackpot, that can be won, between two successive draws is proportional to its preceding value. This growth is best described as a linear function with two parameters a and b reflecting foundational player pool sales and the excitement generated by the current jackpot respectively. A computer simulation considering additional parameters (such as the price of a ticket: s, and the format of the lottery: p) is used to study the statistical features of simulated lotteries covering a vast parameter space. This approach enables us to construct a detailed map of how various parameters influence lottery behaviour by examining the statistical characteristics of the simulated lotteries. The findings emphasize the need for thoughtful pricing strategies in real-world lotteries and suggest that simulations can assist organizers in making informed decisions.*

**Keywords**: econophysics, lottery simulations, computer simulations, complex systems, lottery jackpot dynamics
**JEL Classification**: C55, C61, C63, L83.

*Corresponding author, **István Gere** – istvan.gere@ubbcluj.ro

*Acknowledgements: This work was supported by the project "A better understanding of socio-economic systems using quantitative methods from Physics" funded by the European Union – NextgenerationEU and the Romanian Government, under National Recovery and Resilience Plan for Romania, contract no 760034/23.05.2023, cod PNRR-C9-I8-CF255/29.11.2022, through the Romanian Ministry of Research, Innovation and Digitalization, within Component 9, Investment I8. This paper was presented at the International Conference on Applied Statistics (ICAS) 2023. The authors thank the participants for their useful feedback.

## 1. Introduction

### *1.1. On real lotteries*

Lotteries, as a straightforward form of gambling, enable participants to acquire tickets that serve as entries for a chance to win substantial prizes. It is a widely favoured form of gambling, with global sales estimated at 224.3 billion dollars in the year 2007 (Ariyabuddhiphongs, 2010). When it comes to one of the most widely acknowledged lottery formats, the 6/49 lotto, participants must adhere to the following guideline: after purchasing tickets, they are required to choose 6 numbers from the available pool, ranging from 1 to 49. This personalized number selection constitutes a participant's unique entry into the game, with the option for players to purchase multiple tickets. The main goal of this selected set of numbers is to match those drawn regularly during the lottery proceedings, which usually take place at the central headquarters of the lottery. The jackpot is claimed by the player holding the ticket with precisely matching numbers, while players with tickets showing fewer than six matching numbers (specifically, 5 out of 6, 4 out of 6, and 3 out of 6) receive smaller prizes. These prize amounts are drawn from the funds accumulated during the time interval between two consecutive draws. Subsequently, the jackpot value is typically increased by a portion of the total collected funds in the period between two draws (Baker et al., 2020).

The selection of 6 numbers out of 49 highlights the exceptionally low probability of winning the jackpot. A straightforward combinatorial calculation reveals that there are 13,983,816 ways to fill out a lottery ticket. In contrast, the ways to fill out a ticket that secures a lesser prize are significantly fewer: 1,906,884 for 5 out of 49, 211,876 for 4 out of 49, and 18,424 for 3 out of 49. Consequently, the likelihood of winning the jackpot with a single ticket is exceedingly slim, approximately one in 13 million, while the odds of winning a lesser prize are comparatively higher (Barboianu,2009).

As we described previously, only a fraction of the sales price of a lottery ticket contributes to the main jackpot pool. This practice is rooted in the historical context where lotteries are typically organized by states, considering these games as revenue streams contributing to the state treasury (Willman, 1999). Many contend that lotteries can be seen as a type of regressive taxation, given that lottery players frequently belong to the lower-income strata of society (Pérez and Humphreys, 2012).

It is commonly accepted that approximately 50% of the sales price contributes to the jackpot prize and smaller prizes, forming a substantial portion of the total revenue. The remaining share of the sales is typically allocated to the state, primarily for charitable purposes, and to cover the administrative costs associated with organizing and managing the lottery games (Baker et al., 2020).

In summary, lotteries present themselves as enticing games that draw in numerous players lured by the possibility of winning a fortune, albeit with slim odds. The allure of lotteries is rooted in intricate psychological and sociological factors inherent to the game (Ariyabuddhiphongs, 2010). Additionally, it is worth mentioning that the attraction of lotteries is further heightened by the magnitude of the potential jackpot. Typically, a higher jackpot size correlates with increased ticket sales (Matheson and Grote, 2004).

In the following subsection we will argue that lotteries may be considered as a complex system. Such systems can be comprehended through coarse graining. Our analysis will focus on the governing dynamics, and through the mathematical representation of these dynamics, we aim to develop computer simulations. This approach seeks to comprehend the intricate behaviour of lotteries and identify the fundamental parameters that characterize them.

### *1.2. Lottery as a complex system*

Systems that consist of numerous interacting constituents, adherent to specific rules, are categorized as complex systems. These systems may demonstrate self-organizing principles (Bak, 2013).

In this subsection we assert the complex systems nature of lotteries with the following simple example: As of June 29, 2023, the Romanian lottery's website reported a total prize pool of 1.085.667 RON for Lotto 6/49.The significance of the date in this example lies in the fact that the jackpot was won during the previous draw. Therefore, this case can be considered as an example for baseline sales. Taking into account that a ticket for the Lotto 6/49 costs 7 RON, and typically 50% of the sales contribute to the prize pool, we can calculate that approximately 300.000 individual tickets were sold in just 4 days. Certainly, although there could be variations in the distribution between the funds allocated to the prize pool and the portion designated as tax, we believe that this approximation serves as a valid justification. It reinforces our assertion that, even in the case of lower jackpot values, the significant number of players contributes to classifying the game as a complex system.

In prior research (Gere et al., 2023), we identified the following characteristics that support the idea of lotteries as complex systems:

- Player-Game Interaction: Lotteries emerge from the interplay between players and the established rules.
- Substantial Sales: Lotteries generate significant sales because of the enticing jackpot, despite the low odds of winning.
- Dynamic Jackpot Formation: The jackpot emerges from player contributions and adherence to game rules during each draw.
- Self-Exciting Process: Observable patterns indicate a self-exciting process, where the value of the jackpot depends on its previous value during the preceding draw.

The jackpot time-series offers valuable insights into the dynamics of the lottery, revealing a correlation between jackpot growth and ticket sales. We will utilize this information to develop a simulation approach for describing lotteries. Our simulation will employ a simple toy model, well-suited for computer-based studies.

In this article, we adopt a complex systems perspective, inspired by the field of econophysics. Making our approach distinct, we leverage principles from econophysics, a field that has successfully employed computer simulations for qualitative modelling of wealth distribution (Drăgulescu and Yakovenko, 2000). The computer simulation approach allows us to visualize the evolution of the simulated lottery time-series and explore a portion of the parameter space for potential lotteries, revealing the dynamics of the jackpot prize in time, while allowing us to study its statistical features. Observing and studying the growth dynamics of real lotteries, we define the main parameters that guide the evolution of the jackpot prize. We employ a simplified version of lotteries in our simulations based on the observation, considering a wide range of parameters. In this approach we map the statistical features of possible lotteries, even ones that do not exist in reality. We will present heatmaps of said properties, and look for extremities, parameter sets that for one reason or another lead to dysfunctional lotteries. Our visualisations highlight the dependence of the lottery features on the parameters presenting non-trivial picture of lotteries.

## 2. The case of real lotteries – Growth dynamics

In this section, we will present our observations on the evolutionary dynamics of two real-world lotteries: the Romanian Lotto 6/49 and the international European lottery, Eurojackpot. The jackpot time-series data for the Romanian Lotto 6/49 were obtained from the official website of the Romanian state lotto (www.loto.ro). The Eurojackpot data was sourced from the crowd-funded lottery data collection website (www.beatlottery.co.uk). The collectively referenced data, hosted openly on the Figshare repository, is denoted as (Gere et al., 2023a).

As illustrated in the first row of Figure 1, the dynamics of lotteries exhibit distinct phases. These include periods of growth, during which the jackpot accumulates, followed by a reset triggered when a winner is drawn. On the second and third row of Figure 1 we present the mean and the variance of the jackpot time-series calculated up to each drawing.

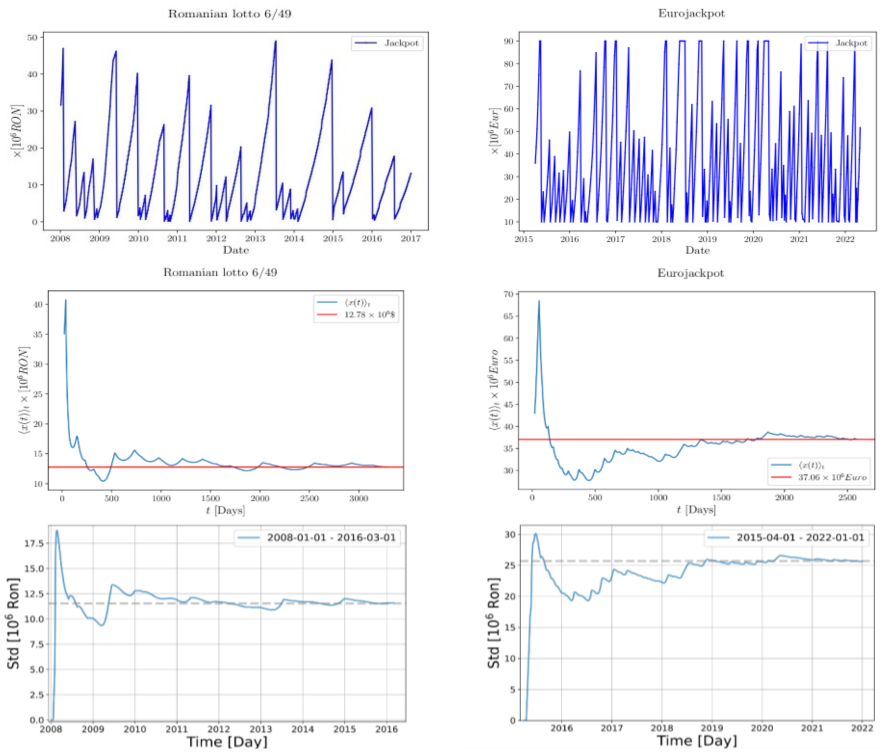


**Figure 1: Our observations regarding the real life lotteries: Romanian lotto 6/49, between the years 2008 and 2017 (left panels), Eurojackpot lotto, between 2015 and 2022 (right panels).**

*Source: Based on the data collected by the authors, freely available in (Gere et al., 2023a)*

Observing these figures, we note that both the mean and the variance converge over time toward a constant value. This observation enables us to consider the lottery time-series as stationary from a statistical standpoint, facilitating the calculation of statistical moments and the identification of the principal growth dynamics of the lotteries.

In describing the relevant growth rates, we will use the mean rescaled jackpot notation: $x/\langle x \rangle$, where $x$ is the value of the jackpot, while $\langle x \rangle$ is the mean of the time-series. This notation allows us simplify our observations, as the mean-scaled jackpot values are smaller and are easier to manage. From the Jackpot values between consecutive draws without winners, the growth rate was empirically determined by computing the average relative increase. Subsequently we use the methodology in (Gere et al., 2023), the calculated growth rate is expressed as a function of the relative jackpot values, averaged within consecutive intervals $B$ of relative jackpot values, as described in (Equation 1.)

$$\mu\left(\langle \frac{x}{\langle x \rangle_t} \rangle_B\right) = \langle \frac{1}{\langle x \rangle_t} \frac{dx}{dt} \rangle_B \tag{1}$$

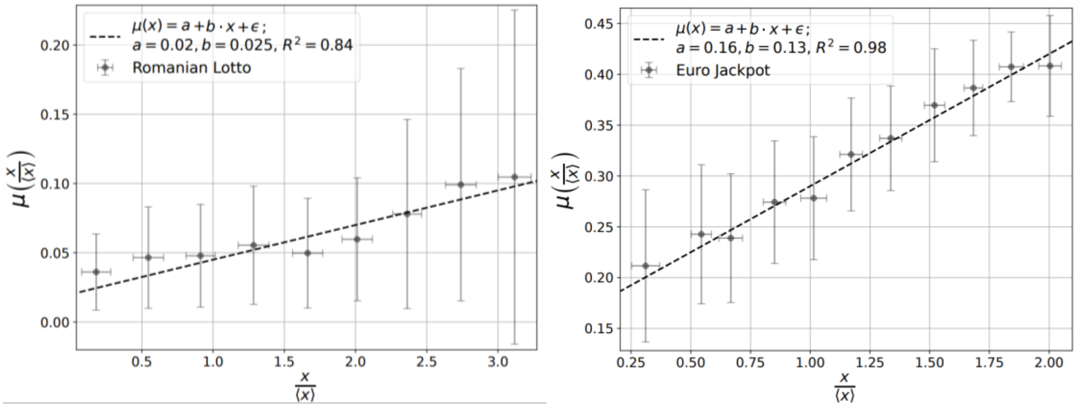


**Figure no. 2: The experimental averaged growth rate calculated from (Equation 1.) for the Romanian Lotto 6/49 (left panel) and Eurojackpot lottery (right panel), and the linear fits.**

*Source: Based on the data collected by the authors, freely available in (Gere et al.,2023a)*

As depicted in Figure 2, the growth rate of the jackpot prize can be effectively characterized by a linear function expressed in the form (Equation 2.) as the high value of $R^2$ (Figure 2) suggests.

$$\mu(x) = a + bx + \varepsilon \quad (2)$$

In addition, ANOVA results further validate the model's statistical significance, with F-test values indicating that the linear functions significantly explain the variance in the averaged rates (for the Romanian Lotto 6/49: $F = 0.251$, $p = 0.0282$ and for the Eurojackpot: $F = 0.263$, $p = 0.0113$), demonstrating significance at the $p < 0.05$ level.

Here *μ(x)* represents the growth rate where *x* is the value of the jackpot while *a* and *b* are the parameters, that characterize the lottery with meaningful economic connotations. The term ε represents the error component, accounting for stochastic variability in the data not captured by the explanatory variables. The inclusion of ε enhances the realism of the linear fits by acknowledging inherent uncertainties in lottery dynamics, such as individual variations and unobserved factors. Such an approach is similar to the regression approach used by economists (Oster, 2004) (Baker et al., 2020) to predict the sales of a lottery as a function of time.

Parameter a is associated with the portion of ticket sales contributing to the jackpot, specifically when a winner is found, resulting in $x = 0$. In this scenario, the jackpot value will have a value around the parameter *a*. It can be considered that parameter *a* is linked to the foundational player pool (those who consistently engage in the game) contributing a fundamental sum to the jackpot pool in each draw. Parameter *b* is associated with the jackpot's present value in comparison to its value following the next draw. It functions as a dimensionless parameter, characterizing how the player pool responds to the increased jackpot value.

As depicted in the mean-scaled notation in Figure 2, Eurojackpot exhibits a larger base player pool, and the players demonstrate heightened responsiveness to the jackpot value. As a consequence, this results in faster growth and shorter intervals between won jackpots, as illustrated in Figure 1, Eurojackpot time-series.

**Table no. 1: The statistical properties of the studied real-world lotteries.**

| Lotto | a/⟨x⟩ | b/⟨x⟩ | Mean Jackpot (⟨x⟩) | Standard deviation (σ) | Average number of draws between two jackpots (⟨Δt⟩) | Mean scaled standard deviation (σ/⟨Δt⟩) | Probability that a single ticket is winning (p) |
|---|---|---|---|---|---|---|---|
| Romanian Lotto 6/49 | 0.02 | 0.025 | 12785933 RON | $11\times10^6$ | 23.74 | 0.86 | $1/(13\times10^6)$ |
| Eurojackpot | 0.16 | 0.13 | 37062674 EUR | $26\times10^6$ | 5.8 | 0.701 | $1/(139\times10^6)$ |

*Source: Based on the data collected by the authors, freely available in (Gere et al., 2023a)*

In Table 1, we present an overview of the statistical features of the studied lotteries. Based on our findings regarding jackpot time-series stationarity, growth rate determination, and parameter interpretation, along with the derived statistical properties for each lottery, we construct our toy model computer simulation, replicating lottery behaviour in the following section.

## 3. The simulation of the lottery dynamics

In the initial phase of constructing our computer model, we outline the simplifications we adopted in its creation. We primarily focus on the continuous growth of the main jackpot, disregarding the portion of sales designated as tax. Furthermore, we assume that the lesser prizes are not paid out from the jackpot pool but rather from the sales accumulated since the last draw. As discussed in the previous section, the jackpot consistently grows between two draws, and this growth is governed by the growth rate in (Equation 2.), characterized by the parameters $a$ and $b$.

We further define the parameter $s$, simulating the price of a single ticket (excluding taxes), allowing us to calculate the number of tickets sold based on the jackpot growth rate $\mu(x)$ given the dimension of this quantity as money. Additionally, we define $p$, representing the probability that a randomly selected ticket is a winner. The final parameter in the model is T, which indicates the number of draws to be simulated. It should be sufficiently large to allow the simulated lotteries to reach a stationary state. In the following section, we outline the algorithm employed for generating simulated lottery jackpot time-series:

1. Define the parameters $a$, $b$, $s$, $p$ and $T$ for the simulation. Throughout all simulations, T is set to 50000 draws. To address computational constraints, we choose to simulate a simplified *4/30* lottery for which $p = 1/27405$.
2. Let $J(t)$ denote the value of the jackpot at the $t$-th drawing.
3. For the initial case at $t = 0$ the jackpot value is set to $J(t=0) = a$, assuming that in this scenario, the sales are generated by the default player pool.
4. For the consecutive draws $t > 0$:
    a. Calculate the total number of tickets $n$ sold from the price parameter $s$ and the value of the jackpot growth $\mu(J(t-1))$ representing the sales since the previous draw $t$ as (Equation 3.):

$$n = \frac{\mu(J(t-1))}{s} \quad (3)$$

b. Verify if there is a winner by generating *n* uniform random numbers between *0* and *1* (denoted as *r*), representing the tickets sold. If $r < p$, a winner is found; otherwise, there is none:
   i. If there is no winning ticket, the value of the jackpot at the current time, t, will be equal to:
   $$J(t) = J(t-1) + \mu(J(t-1)) \quad (4)$$
   ii. If there is a winning ticket, the value of the jackpot resets to its default value as described in step 3: *J(t) = a*
   iii. The timestep t is incremented with one, *t = t + 1.*

c. Repeat the steps under point b) until *t* reaches the desired limit *T = 50000.* It is essential to note that the parameters of *a* and *b* are varied with a ±*5%* of their values at random during the simulation at each step, to account for possible fluctuations.

Using the algorithm defined above, we generated time-series that exhibit similar properties to the ones observed in the experimental data. This analysis aimed to reveal the impact of parameter variations on the simulated lottery's behaviour, offering insights into the dynamics of these systems. The parameter space consists of 5500 parameter combinations of the values defined bellow:

- *a =[0.01 , 0.022, 0.035, 0.048, 0.06 , 0.072, 0.085, 0.097, 0.11 , 0.122, 0.135, 0.148, 0.16 , 0.172, 0.185, 0.197, 0.21 , 0.222, 0.235, 0.248, 0.26 , 0.272, 0.285, 0.298, 0.31 ]*
- *b = [0.005, 0.031, 0.057, 0.083, 0.109, 0.135, 0.161, 0.187, 0.213, 0.239, 0.266, 0.292, 0.318, 0.344, 0.37 , 0.396, 0.422, 0.448, 0.474, 0.5 ]*
- *s =[1.0e-07, 2.7e-07, 1.0e-06, 5.0e-06, 1.0e-05, 5.0e-05, 1.0e-04, 5.0e-04, 1.0e-03, 5.0e-03]*

## 4. The interpretation of the results

The vastness of the parameter space makes it challenging to visualize its effects comprehensively. Our focus will be on examining the impacts of parameters on the mean (⟨*J*⟩), standard deviation ($\sigma$) and the average number of draws (⟨Δt⟩) between jackpot wins and exploring the probability distributions of the jackpots. To see which simulated lotteries are realistic first we will study the average number of draws (⟨Δt⟩) between jackpot win on Figure 3.

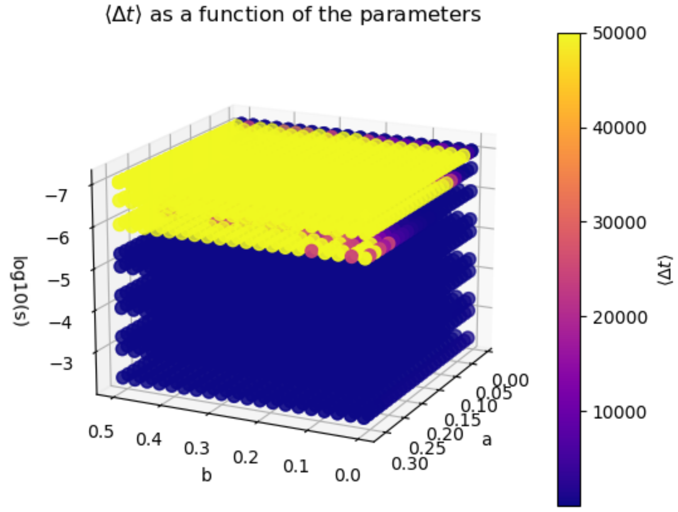


**Figure no. 3: The heatmap of the average number of draws ⟨Δt⟩ between jackpots won.**

*Source: Based on the computer simulations of the toy model.*

Observing Figure 3, it becomes apparent that certain parameter combinations can lead to the lottery becoming 'stuck.' In some scenarios, the low *s* price parameter and a relatively high *a* base sales parameter result in the lottery being won at every draw. Conversely, due to random variations introduced to the growth parameters, the lottery jackpot may only grow once or twice sporadically before being won. It is crucial to note that such instances do not typically occur in real lotteries because game pricing is usually adjusted based on the base number of players to prevent such situations. However, in reality, such cases are not observed. Theoretically, the guiding dynamics of lotteries allow for the existence of such scenarios. On Figure 4, we present the same heatmap filtered, for sequences where the average time between jackpots won is smaller than $\langle \Delta t \rangle < 1000$.

Based on the observations we can see from Figure 4 that although for some parameter pairs our simulated *4/30* lottery realises dysfunctional lotteries, many of the remaining parameter pairs after filtering convey a possible realistic lottery.

Given the limitation of presenting all statistical features on three-dimensional heatmaps, the subsequent sections of this paper will focus on specific segments from this heatmap. We will explore the impact of the *a* and *b* parameters on relevant statistical properties while keeping the s parameter constant. Additionally, we will investigate the influence of different *s* parameters with fixed *a* and *b* values.

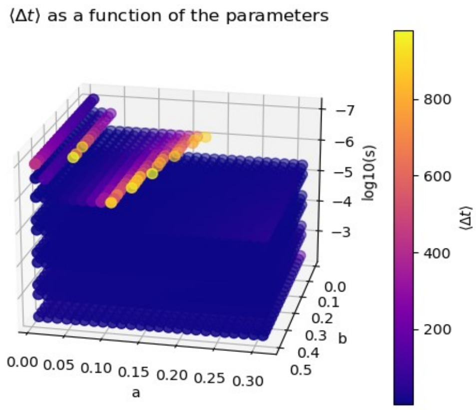


**Figure no. 4: The heatmap of the average number of draws between jackpots won, filtered as only values for which the ⟨Δt⟩ < 1000 are shown.**

*Source: Based on the computer simulations of the toy model.*

On Figure 5, the relationship between the average number of draws between wins for a single (fixed) *s* price parameter and the growth parameters is represented as a heatmap on the left panel. The plot on the right panel showcases the dependence of the average number of draws between wins (⟨Δt⟩) as a function of parameter s for multiple growth rate parameter pairs. Note that the first few datapoints on the left panel represent the 'stuck' lotteries we mentioned before. The right panel clearly shows an increase in the player pool parameter *a* and the excitation parameters *b* leads to a rapid decrease in the average time between draws, while the left panel presents that the average number of draws (⟨Δt⟩) between wins scales as a power law with the value of the price parameter *s*.

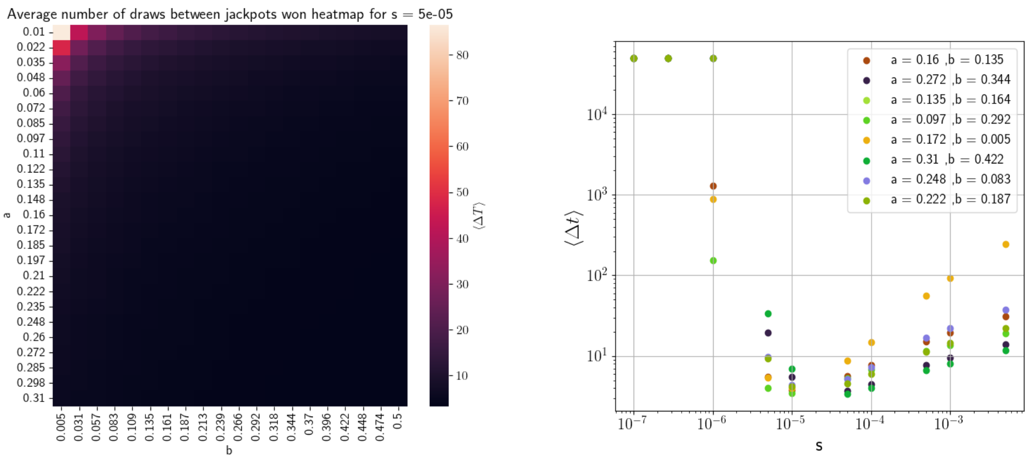


**Figure no. 5: The heatmap of (⟨Δt⟩) as function of the growth parameters (left panel). The (⟨Δt⟩) as a function of the s parameter for multiple growth rate parameters (right panel).**

*Source: Based on the computer simulations of the toy model.*

Bellow, in Figures 6 and 7, we present in a similar manner as described above the heatmaps that illustrate the average jackpot (⟨J⟩) and the standard deviation (σ) of the time-series concerning the growth parameters $a$ and $b$ for a fixed $s$ on the left panel. Additionally, the dependence of these statistical measures on the $s$ parameter is explored for randomly selected growth parameter pairs on the right panels.

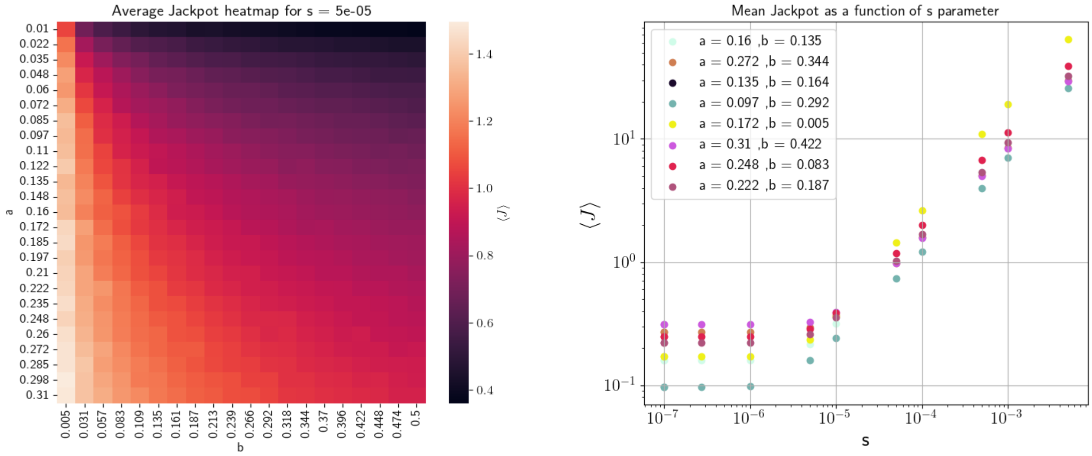


**Figure no. 6: The heatmap of the (⟨J⟩) average jackpot of the time-series as function of the growth parameters (left panel). The (⟨J⟩) as a function of the s parameter (right panel).**

*Source: Based on the computer simulations of the toy model.*

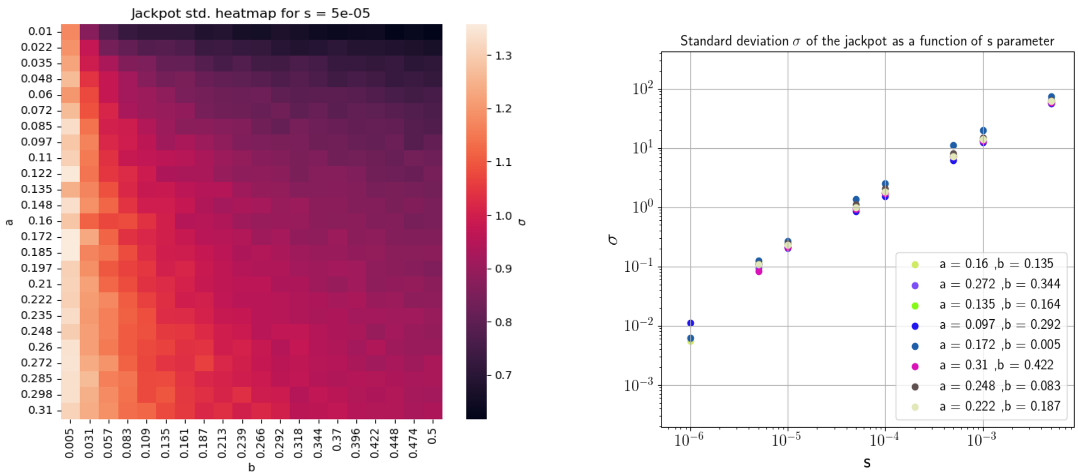


**Figure no. 7: The heatmap of the standard deviation (σ) of the jackpot time-series as function of the growth parameters (left panel). The (σ) as a function of the s parameter (right panel).**

*Source: Based on the computer simulations of the toy model.*

Our observations for the Figures 6 and 7 are the following: from the left panels we can clearly observe, that the dependence of the mean (⟨J⟩) and the standard deviation (σ) for a fixed $s$ parameters are similar. We can see that the increase in the player pool parameter $a$ leads to and increased average jackpot and standard deviation, while the excitation parameters $b$ has a dampening effect, its increase decreases both metrics, an outcome that is not trivial. In the right panels, a power-law like increase of the mean jackpot (⟨J⟩) as a function of $s$ is observed. This observation holds for sufficiently large s parameters, ensuring that the simulated lotteries avoid being 'stuck'. The right panel on Figure 7 presents the standard deviation (σ) as a function of the *s* price parameter, specifically for s values where the lottery is no longer 'stuck', and here we once again observe a power-law like increase. As these results show, our model is capable of generating lotteries with diverse statistical features.

Last we present as examples of simulated lotteries the probability distribution of jackpots *ρ(x)* for some of the simulated lotteries along with time-series and some statistical features. In Figure 8, probability distributions of simulated lotteries are presented alongside the probability distribution of the Romanian Lotto 6/49. Despite differences in format between the simulated and real lotteries, the real-world lottery serves as a useful reference for comparison.

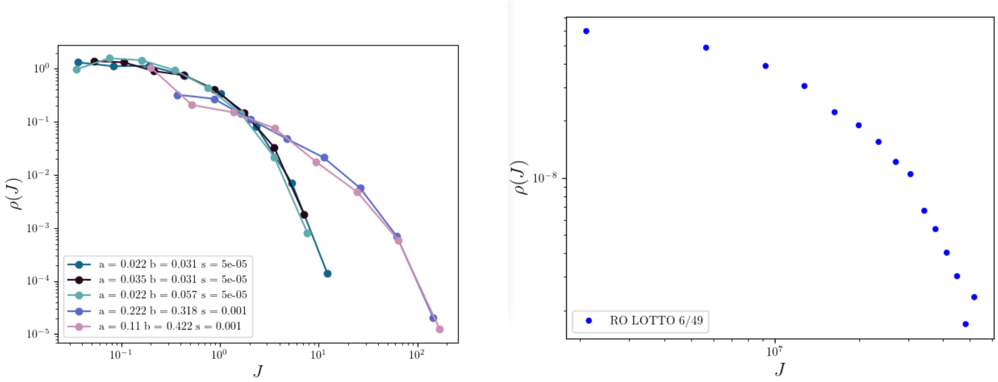


**Figure no. 8: The probability distribution of the jackpots for the simulated lotteries along with the parameters (left panel). A reference of the Romanian Lotto 6/49 (right panel).**

*Source: Based on the computer simulations of the toy model and on the data collected by the authors, freely available in (Gere et al.,2023)*

The time series of the simulated lotteries can be visualized similarly to how we presented the time series of real-world lotteries in Figure 1, and we present them accordingly in Figure 8.

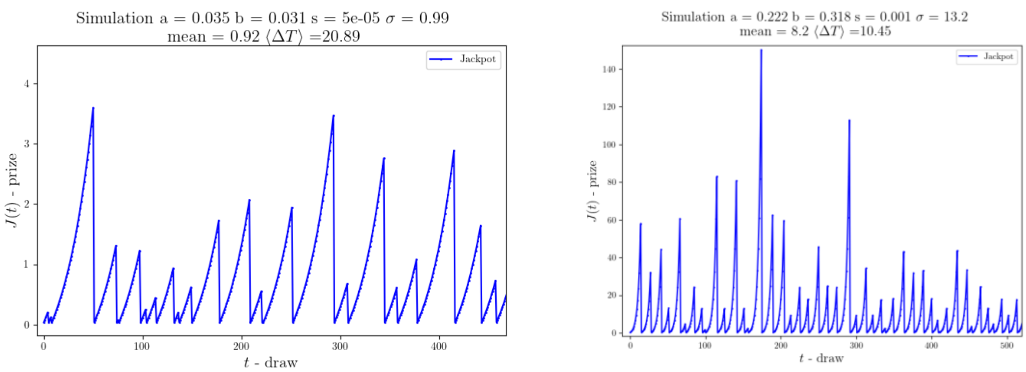


**Figure no. 9: The time series of two examples of simulated lotteries selected from the ones presented on Figure 8.**

*Source: Based on the computer simulations of the toy model.*

As depicted in Figure 9, the simulated lotteries showcase dynamics characterized by phases of growth followed by a reset to the default value. We can also clearly observe the effects of the parameters as the time-series presented on the left panel of Figure 9 has lover values for each of its parameters *a, b* and *s*, than the one presented on the right panel, leading to a slower dynamics with a larger average time between jackpots won. The statistical features of the simulated lotteries, along with ($\langle J_{win} \rangle$) the average value of the jackpot won, and (*Max(J)*) the maximal jackpot in the time-series, for which we presented the probability distribution of the jackpots on Figure 8, and one additional example are presented in Table 2.

**Table no. 2: The statistical properties for selected examples of simulated lotteries.**

| Lotto | a | B | s | ⟨J⟩ | Max(J) | ⟨ Δt ⟩ | $\langle J_{win} \rangle$ | σ |
|---|---|---|---|---|---|---|---|---|
| Sim. #1 | 0.022 | 0.031 | 0.0005 | 0.82 | 17.14 | 27.61 | 1.39 | 0.97 |
| Sim. #2 | 0.035 | 0.031 | 0.0005 | 0.922 | 9.55 | 20.88 | 1.41 | 0.99 |
| Sim. #3 | 0.022 | 0.057 | 0.0005 | 0.70 | 10.51 | 21.12 | 1.36 | 0.71 |
| Sim. #4 | 0.022 | 0.031 | 0.001 | 8.19 | 203. | 10.44 | 24.06 | 13.2 |
| Sim. #5 | 0.11 | 0.422 | 0.001 | 6.66 | 244.53 | 10.51 | 23.07 | 12.07 |
| Sim. #6 | 0.11 | 0.318 | 0.0005 | 4.13 | 100.78 | 10.48 | 12.15 | 6.63 |

*Source: Based on the computer simulations of the toy model.*

As highlighted in Table 2 and illustrated by Figures 8 and 9, our selected examples of simulated lotteries exhibit diverse statistical features, demonstrating the impact of different parameter sets on the statistical properties of potential lotteries. This approach allows us to study the effects of individual parameters by querying the results based on specific filters.

## 5. Conclusion

In this paper, we adopt a straightforward econophysics approach to examine the time-series statistics of lotteries. Treating lotteries as complex systems, we studied the guiding dynamics of real-world lotteries. Using the methodology of our previous work (Gere et al., 2023) our observations revealed that the organizing principle of lotteries is a growth in the jackpot from one draw the next that is proportional to the jackpot's value in the preceding draw. The growth is characterised by a linear function in the form of (Equation 2.) with two main parameters, *a* and *b*. In this approach *a* is connected to the to the sales generated by the foundational player pool, while *b* is a dimensionless parameter that describes the sales generated by the excitation caused by the current value of the jackpot in the player pool.

Based on these observations, we propose a computer simulation that introduces three additional parameters: *s*, representing the unit price of a ticket used to calculate the number of tickets sold; *T*, the number of draws simulated; and *p*, the probability of a ticket being a winner (calculated based on the lottery format). Due to computation constraints in our simulations we study a lotteries for which $p = 1/27405$ corresponding to a lottery in the format 4/30. Using this simulation we map the statistical properties of a space of 5500 possible parameter sets. The strength of the computer simulations over the analytical description described in (Gere et al.,

2023) lies in the fact that it is capable in recreating time-series of lotteries allowing us to study the mean time spent between the jackpots being won.

Three dimensional heatmaps are created regarding the average number of draws elapsed between two jackpots won ($\langle\Delta t\rangle$) as a first step to understand the effects of the parameters. We observe, that for certain sets of parameters the lottery 'breaks down' meaning that game with such features would not work in reality. In this cases the initial sales defined by parameter *a* coupled with the low ticket prices defined by *s* create such a condition that so much tickets are sold at every first step that the lottery will always get won. Paradoxically, these lotteries are present in the generated time series with $\langle\Delta t\rangle = 50000$, as no peaks are found in the data to calculate the elapsed times between jackpots won. Alternatively, sporadically, it grows once or twice and then gets won, due to our methods of looking for peaks in the data and assigning these to the cases of the jackpot being won, such cases will also appear in our data with a high ($\langle\Delta t\rangle$) value. Using the condition that $\langle\Delta t\rangle < 1000$ is a good indicator that the simulated lottery did not fall into the case of dysfunctional lotteries. In our simulation there are 3811 parameter sets for *a, b* and *s* for which this condition is true. This simple observation highlights the importance of carefully selecting the pricing for real-world lotteries, taking into account the default sales generated by the lottery.

Studying a section for a fixed *s* parameter of the three dimensional heatmap we observed that the value of ($\langle\Delta t\rangle$) decreases quickly with the increase of parameters *a* and *b*. In similar heatmaps for the mean ($\langle J\rangle$) and the standard deviation ($\sigma$) of the studied time series we observe the non-trivial fact that these measure do not change in the same manner as ($\langle\Delta t\rangle$). The mean and the standard deviation increases with the value of parameter *a* while decreases in a non-trivial way with the increase of the *b* excitation parameter. On plots with fixed *a* and *b* values we studied the effect of parameter *s* on the statistical properties of the time-series. In the case of the simulated lotteries not falling into the 'broken down' category, the relevant statistical properties increased in a power-law fashion with the rise of the price parameter *s*.

We also presented figures with the probability distribution of jackpot for the simulated lotteries and the value of the statistical features for a selected few simulations. The time-series were also presented for two simulated lotteries.

In conclusion, we believe that simulations, as presented in this article, can offer valuable insights into the statistical features of lotteries. By evaluating real-world lotteries, this approach can assist lottery organizers in making informed decisions, such as adjusting ticket prices (affecting parameter *s*) or refining advertising strategies (affecting parameter *b*). Jackpot evolution simulations can assist in validating that the lottery operates within regulatory constraints.

Further studies could be dedicated to the study of the non-trivial power-law like scaling observed between the price parameter *s*, and the statistical features of lotteries, and to the decrease of the mean jackpot and the standard deviation with the increase of the *b* excitation parameter.

## References


Ariyabuddhiphongs, V., 2010. Lottery gambling: A review. *Journal of Gambling Studies*, 27, 15–33. https://doi.org/10.1007/s10899-010-9194-0.

Bak, P., 2013. *How nature works: the science of self-organized criticality*. Springer Science & Business Media.

Baker, R., Forrest, D. and Pérez, L., 2020. Modelling demand for lotto using a novel method of correcting for endogeneity. *Economic modelling*, 84, 302–308. https://doi.org/10.1016/j.econmod.2019.04.021

Barboianu, C., 2009. *The Mathematics of lottery: Odds, Combinations, Systems*, INFAROM Publishing.

Drăgulescu, A. and Yakovenko, V. M., 2000. Statistical mechanics of money. *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(4), 723–729. https://doi.org/10.1007/s100510070114

Gere, I., Kelemen, S., Néda, Z. and Biró, T.S., 2023. Jackpot statistics, a physicist's approach. *arXiv (Cornell University) [Preprint]*. https://doi.org/10.48550/arxiv.2311.04826

Gere, I., Kelemen, S., Néda, Z. and Biró, T.S., 2023a. Lottery Jackpot Time-Series. [online]. *Figshare repository*, https://doi.org/10.6084/m9.figshare.24310669

Matheson, V.A., and Grote, K., 2004. Lotto fever: do lottery players act rationally around large jackpots?, *Economics Letters*, 83(2), 233–237. https://doi.org/10.1016/j.econlet.2003.11.010.

Oster, E., 2004. Are All Lotteries Regressive? Evidence from the Powerball. *National Tax Journal*, 57(2), 179–187. http://www.jstor.org/stable/41790148

Pérez, L. and Humphreys, B.R., 2012. The ‘Who and Why’ of Lottery: Empirical Highlights from the Seminal Economic Literature. *Journal of Economic Surveys*, 27 (5), 915–940. https://doi.org/10.1111/j.1467-6419.2012.00732.x.

Willman, G., 1999. *The history of lotteries*, Unpublished Manuscript, Stanford University, California, USA, http://willmann.com/~gerald/history.pdf